1

2

3

4   UNITED STATES DISTRICT COURT

5   NORTHERN DISTRICT OF CALIFORNIA

6   EUREKA DIVISION

7

8   PAUL TREMBLAY, et al.,   Case No. 23-cv-03223-AMO (RMI)

9   Plaintiffs,

10   v.   **ORDER RE: DISCOVERY DISPUTE
                                                 LETTER BRIEFS**

11   OPENAI, INC., et al.,   Re: Dkt. Nos. 298, 337, 340, 352

12   Defendants.

13

14   Now pending before the court are a number of discovery dispute letter briefs (dkts. 298,

15   337, 340, 352). Pursuant to Federal Rule of Civil Procedure 78(b) and Civil Local Rule 7-1(b), the

16   court finds that these matters are suitable for disposition without oral argument. As to the disputes

17   presented, the court rules as follows.

18   ***Plaintiffs' Proposed Additional Search Terms (dkt. 298)*:**

19   The court previously denied Plaintiffs' suggestion that the requesting Party have input in

20   determining the search terms "because it raised the specter of too many future delays and disputes

21   over methodology and search term formulation." *See* Order of July 31, 2024 (dkt. 166) at 3.

22   Several months later, while conceding that OpenAI had not excluded documents containing the

23   term "torrent" (and that OpenAI had indeed produced some documents responsive to an RFP

24   pertaining to the use of torrents), Plaintiffs nevertheless alluded to gaps in the production and once

25   again sought to inject themselves into the ESI search-term formulation process. *See* Ltr. Br. (dkt.

26   242) at 3. Finding that Plaintiffs had failed to show why the production was insufficient, the court

27   denied Plaintiffs' suggestion that they should have input into the formulation of search terms. *See*

28   Order of January 13, 2025 (dkt. 247) at 2-3.

United States District Court
Northern District of California

1   It appears that Plaintiffs now seek to litigate that issue for a third time, and by doing so, it

2 appears that the objective of the court's initial ruling (*i.e.*, to prevent delays and the stalling of the

3 discovery process stemming from endless discovery disputes over methodology and search-term

4 formulation) has been nevertheless frustrated. While OpenAI submits that Plaintiffs have

5 identified no deficiencies in its productions (*see* Ltr. Br. (dkt. 298) at 5) – Plaintiffs essentially

6 respond to the effect that they "by definition know little about documents OpenAI hasn't

7 produced." *Id*. at 2. Plaintiffs note that their review of "OpenAI's productions and search terms,

8 identified areas where the terms and connectors appeared too narrow or didn't capture lingo used

9 by OpenAI's own employees, and proposed additional search terms." *Id*. Plaintiffs add that their

10 proposed eight search "strings" are not burdensome because they would yield only 345,000

11 documents, which "[w]ith technology assistance, e-discovery vendor DISCO estimates a 345,000-

12 document review takes 2-3 weeks, [from] upload through production." *Id*. at 4.

13   OpenAI responds by noting that Plaintiffs' request now not only seeks the addition of

14 "numerous torrent-related search terms, but also [] hundreds of more terms packed into eight

15 search strings, including nonsensically overbroad terms like (ChatGPT AND "be doing"),

16 (memori* AND data), and (seed*) . . . [which] [b]ased on review time to date, Plaintiffs' terms

17 would add over 9,000 hours of attorney review time with no showing that OpenAI's productions

18 are deficient." *Id*. at 4. As to proportionality, OpenAI submits that "Plaintiffs purport to propose

19 only 8 search 'strings,' but those search strings consist of 362 search terms, hitting on an

20 additional 345,000 documents beyond the over 640,000 documents that OpenAI has already

21 agreed to review." *Id*. at 6. OpenAI submits that requiting it to introduce hundreds of thousands of

22 non-responsive documents into its review queue will only prolong fact discovery while failing to

23 add to the substance of Plaintiffs' case. *Id*. at 6. The court agrees with OpenAI's assessment here.

24 Without repeating the details of the Parties' disputes as to the numerous search terms included in

25 Plaintiffs' proposed search strings – of which, string numbers 1 and 4 were set forth by OpenAI as

26 exemplars (*see id*. at 5-6) – the court cannot avoid the conclusion that Plaintiffs are trying to

27 relitigate an issue they have already lost (*i.e.*, that they should be involved in the search term

28 formulation process without making a clear showing of prejudice stemming from gaps or

1   deficiencies in the search terms disclosed by the producing Party). As a result, Plaintiffs' request

2   for an order directing OpenAI to deploy its proposed search strings is **DENIED**.

3   ***OpenAI's Request for Discovery from Plaintiffs' Agents, Assistants, etc. (dkt. 337)*:**

4   　　　　OpenAI submits that the twelve Plaintiffs in this action have not adequately searched for or

5   produced "documents from the files of those who work under their direction and on their behalf—

6   their literary agents, ghostwriters, loan-out companies, assistants, and production companies

7   (hereafter, 'Agents')." *See* Ltr. Br. (dkt. 337) at 2. As OpenAI sees it, two disputes exist in this

8   regard: "First, although Plaintiffs do not dispute that the Agents' files are within their 'control,'

9   the parties disagree as to whether Plaintiffs have conducted a reasonable search of those files.

10   Second, the parties disagree as to whether Plaintiffs' publishers' files are within Plaintiffs'

11   'control.'" *Id*.

12   　　　　OpenAI seeks certain information from Plaintiffs' literary and other agents that negotiated

13   publishing and licensing agreements for their works – OpenAI submits that "[d]ocuments relating

14   to such negotiations are relevant to (1) ownership of the works, (2) the purpose of the works, and

15   (3) the market for the works and any alleged harm thereto. Yet, for example, Snyder's and

16   Lippman's productions do not contain any responsive documents regarding the negotiations for

17   their domestic publishing agreements. And Coates has not produced any documents from his

18   literary agent, his film and TV agent, or his manager . . ." *Id*. OpenAI also seeks more discovery

19   from the loan-out companies that are the alleged owners of some of the asserted works, from a

20   number of Plaintiffs' assistants, and from production companies that Plaintiffs control. *Id*. at 2-3.

21   OpenAI then concludes by stating that, "[g]iven the April 28 fact discovery cut-off, the Court

22   should compel all Plaintiffs to conduct a reasonable inquiry for responsive documents from the

23   Agents' files and to produce them by February 28." Plaintiffs respond to the effect that they "have

24   already conducted diligent searches for potentially responsive documents from a variety of data

25   sources and agents, as required by the Federal Rules and the Parties' ESI Protocol [and] [w]hat

26   OpenAI seeks now are documents from individuals who do not possess responsive documents,

27   [or] from individuals from which Plaintiffs have no legal right to obtain documents, or both." *Id*.

28   at 4.

United States District Court
Northern District of California

1   The court finds that this dispute is representative of the Parties talking past one another –

2   that it is the manifestation of poor meet-and-confer efforts – and, that the "dispute" evades judicial

3   resolution as it has been presented. Thus, the Parties are **ORDERED** to meet-and-confer forthwith

4   regarding OpenAI's assertions regarding the facet of their dispute regarding documents from all of

5   Plaintiffs' Agents (not including the publishers, which will be discussed below). Following which,

6   seeing as no genuine dispute about relevance of proportionality has been presented (*see id*. at 3-6),

7   Plaintiffs are **ORDERED** to either produce the requested materials, or to certify to Defendants

8   that they have searched for the documents or information in question and that no such documents

9   or information were found that have not already been (or will be) produced. Thereafter, unless

10   OpenAI can articulate a reason why the court should take further action in this regard (*i.e.*, by

11   making a showing that Plaintiffs' certifications are demonstrably untrue), the matter will be

12   concluded and OpenAI will have to accept the productions and certifications it has received.

13   As to documents and information that OpenAI seeks from Plaintiffs' publishers, OpenAI

14   submits that "Plaintiffs do not dispute the existence of relevant discovery in their publishers'

15   possession. Their only excuse for not producing it is that they do not control their publishers. But

16   each Plaintiff's publishing agreement contains provisions entitling them to obtain from the

17   publisher highly relevant and responsive information." *Id*. at 4. Unfortunately, it once again

18   appears that the Parties are practically speaking different languages and that they are arguing past

19   one another because Plaintiffs respond to the effect that they "do not have a right to receive the

20   documents requested by OpenAI from their publishers [and that] OpenAI misconstrues and

21   oversimplifies the extent of Plaintiffs' rights under the publishing agreements." *Id*. at 6. Plaintiffs

22   add that they "have already produced the documents they are entitled to under their agreements—

23   they have produced sales and royalty statements that Plaintiffs have received from their publishers.

24   But OpenAI demands more, for example, market analyses of the works, but those documents are

25   simply not within the scope of documents that Plaintiffs have a right to obtain, as Plaintiffs have

26   explained to OpenAI. OpenAI has the same rights to the other documents it seeks from Plaintiffs'

27   publishers as Plaintiffs do through Rule 45 subpoenas." *Id*.

28   This dispute too seems to be the product of the Parties seemingly taking turns to simply

4

United States District Court
Northern District of California

1   conjure up something about which to fight. The court expects counsel to resolve this sort of

2   dispute without the need for court intervention. Thus, rather than laying such a dispute at the

3   court's doorstep – if OpenAI believes that Plaintiffs' publishers are in possession of relevant

4   information which Plaintiffs' assert is beyond the scope of their possessory rights under the

5   relevant agreement(s), then OpenAI should – as Plaintiffs have suggested – seek to obtain those

6   documents via third-party discovery practice through a Rule 45 subpoena. Accordingly, as to this

7   facet of the Parties' dispute, OpenAI's request to compel relevant material from Plaintiffs'

8   publishers that goes beyond the scope of what each of those agreements allows Plaintiffs to obtain

9   is **DENIED**. As to any relevant material that falls within the scope of what Plaintiffs may obtain

10 from their publishers under the relevant agreements – to the extent not already produced,

11 OpenAI's request to compel the production of that material is **GRANTED**. If any dispute shall

12 arise as to relevance or proportionality, the Parties are **ORDERED** to meet and confer such as to

13 resolve or substantially narrow that dispute so that it can be articulated in a clear and cohesive

14 manner that would lend itself to a ruling which would not require the expenditure of substantial

15 effort in attempting to square the Parties' arguments with one another.

16 ***Plaintiffs' Request to Identify File Names of Withheld Documents in Privilege Logs (dkt. 340)***:

17   Plaintiffs submit that file names should be included on the Parties' privilege logs because

18 file names are the type of information that will enable Plaintiffs to assess a claim of privilege

19 without revealing the privileged information itself. *See*d Ltr. Br. (dkt. 340) at 2. OpenAI claims

20 that including the file names, at this juncture (while narrating Plaintiffs' reported changes of

21 position as to this issue) is unduly burdensome because it "will result in another round of privilege

22 review and redactions." *Id*. at 4-5. OpenAI adds that "[n]ow, for the first time, following months

23 of negotiations over the contents of the parties' logs, Plaintiffs contend that OpenAI's logged

24 subject matter descriptions are somehow 'uninformative' . . . [a]s such, requiring OpenAI to

25 undertake an additional burdensome privilege review and redaction process for several hundred or

26 more file names is entirely needless. And contrary to Plaintiffs' assertions, file names may be the

27 subject of privilege review and redactions, as they will undoubtedly be here." *Id*. at 5. Plaintiffs

28 counter to the effect that they take issue with OpenAI's claim that Plaintiffs have wasted the

1   parties' time by constantly changing positions, because that "is a misrepresentation of the email

2   correspondence the parties exchanged and met and conferred on during the past month which was

3   done in good faith to come to an agreement on the scope and cut-off date for logging privileged

4   documents [and] Defendants fail to point out [that] they changed their position several times as

5   well which is ordinary in the course of negotiations." *Id*. at 4.

6   The court finds this sort of back-and-forth to constitute yet another manifestation of the

7   Parties' drumming up a discovery fight about a subject that should have been worked out by the

8   Parties themselves without the need for court intervention. The dispute itself is insubstantial – the

9   essence of which boils down to little more than petty finger-pointing about whose time was wasted

10   by whom. Given Plaintiffs' statement to the effect that "[t]he need for including filenames is

11   especially necessary in light of OpenAI's inadequate privilege logging practices to date," it is

12   unclear why Plaintiffs waited until February 21, 2025, to first raise this matter. Nor have Plaintiffs

13   presented a specific and persuasive case that they have been prejudiced by the omission of the files

14   names they now claim should have been part of OpenAI's privilege logs all along. *See id*. at 2-4.

15   Instead, even putting aside the sheer volume of seemingly endless discovery disputes in this case,

16   the timing and nature of this dispute (in conjunction with the other disputes that are the subject of

17   this order) gives rise to the impression that the Parties may be using the discovery process, at least

18   in part, to engage one another in a warfare of sorts, that is, by attempting to fashion certain

19   discovery disputes in order to harass one another's flanks as might be the case during a game of

20   chess or on a medieval battlefield. Regardless of whether or not that impression holds true – at

21   bottom, because the court finds that Plaintiffs could have, and indeed should have, presented any

22   issue they may have had with *the form* of OpenAI's privilege log at a much earlier juncture during

23   the discovery process, Plaintiffs' request for the court to intervene at this late stage in the

24   discovery process regarding so ancillary a matter as a minute detail as to the form of OpenAI's

25   privilege log is **DENIED**.

26   //

27   //

28   //

United States District Court
Northern District of California

6

1   ***OpenAI's Requests for Discovery from the Kadrey v. Meta Platforms Case (dkt. 352)*:**

2   OpenAI relates that it seeks discovery from Plaintiffs in a separate, but reportedly similar,

3   action that Plaintiffs have filed against Meta Platforms, Inc., in *Kadrey v. Meta Platforms, Inc.*,

4   No. 3:23-cv-03417 (N.D. Cal.) – through which OpenAI submits that Plaintiffs have asserted

5   copyright infringement claims that are materially identical to the claims in this case. *See* Ltr. Br.

6   (dkt. 352) at 2. According to OpenAI, in both cases, Plaintiffs have alleged Meta and OpenAI

7   trained their large language models using Plaintiffs' books which deprived Plaintiffs of payments

8   for the use of their asserted works as training data. *Id*. Given what OpenAI considers to be a

9   substantial overlap between the legal and factual issues in the two cases, it wants discovery from

10   Plaintiffs' involvement in the *Kadrey* case along the following lines: "(a) Plaintiffs' responses to

11   interrogatories and requests for admission ("RFAs"), Ex. A (RFPs 88/89/90); (b) expert witness

12   materials disclosed by or on behalf of Plaintiffs, Ex. B (RFPs 86/87/88); and (c) deposition

13   transcripts and exhibits for Plaintiffs' expert witnesses, Ex. C (RFPs 87/88/89)." *Id*. OpenAI then

14   states that while "Plaintiffs concede that cross-production is appropriate 'to the extent that the

15   facts overlap,' [h]ere the overlap is extensive and includes: (1) whether Plaintiffs own the

16   copyrighted works they assert in both cases (relevant to Plaintiffs' infringement claim); (2)

17   whether a relevant market for Plaintiffs' works has been harmed by the conduct at issue (relevant

18   to the defendants' primary defense in both cases – fair use []); and (3) what caused the harm

19   Plaintiffs claim to have suffered, which, for example, Plaintiff Hwang admitted is the 'same harm'

20   in both cases." *Id*. at 4. OpenAI repeatedly characterizes its requests as "specific" and "targeted."

21   *See id*. at 2-4.

22   Right off the bat, the undersigned struggles to comprehend the usefulness of OpenAI's

23   interest in seeking information from the *Kadrey* case as to whether or not Plaintiffs actually own

24   the copyrighted works they have asserted in both cases. Unless there is a specific issue as to

25   ownership (and none has been articulated in this letter brief by OpenAI) the undersigned sees no

26   material advancement of OpenAI's defense in this case by putting Plaintiffs to the burden of

27   producing material from the *Kadrey* case that simply confirms what OpenAI already knows (or

28   will know) to be true in this case – that is, that Plaintiffs are the owners of the works at issue in

1   this case. The court would view the issue differently if, on the other hand, OpenAI had articulated

2   a basis to conclude that ownership of the copyrighted works at the heart of this case was a

3   contested issue here, and that discovery about that topic from *Kadrey* might be logically related

4   *and* useful to the resolution of that issue or to OpenAI's defense in general. However, no such

5   showing having been made, the court sees no reason to put Plaintiffs to the burden of producing

6   discovery from *Kadrey* related to their ownership of the works at issue in this case. If OpenAI can

7   better articulate a basis for seeking that information, the Parties should meet and confer in an

8   effort to resolve the matter informally, but if such efforts fail, OpenAI can re-present its request

9   for a court order on the basis of a better-articulated foundation. In the meantime, as to OpenAI's

10   request to compel Plaintiffs to produce information from *Kadrey* related to their ownership of the

11   works at issue in this case, that request is **DENIED**.

12   As to the other two broad categories of information that OpenAI seeks from the discovery

13   process in *Kadrey* (whether a relevant market for Plaintiffs' works has been harmed by Meta's

14   conduct, and what caused the harm that Plaintiffs claim to have suffered as a result of Meta's

15   conduct) – having reviewed the Parties' statements and their arguments, the court is unable to see

16   the usefulness of this discovery in the context of OpenAI's defense in *this* case. For example, as to

17   the interrogatory and request for admission responses OpenAI seeks from *Kadrey*, it asserts that

18   the "discovery sought, *inter alia*, details on Plaintiffs' claims of ownership of their asserted works,

19   Plaintiffs' positions on the markets for those works, and admissions regarding whether Plaintiffs

20   have licensed those works to train LLMs . . .[and that] Plaintiffs' positions on those issues – where

21   the same works are asserted, the same alleged market and market harm is claimed, and the same

22   damages are asserted – are directly relevant to Plaintiffs' claims of copyright infringement and

23   OpenAI's fair use defense in this case." *Id*. What the court struggles to understand is why OpenAI

24   needs that information from *Kadrey*, why it would not be duplicative of the information is has

25   solicited or can solicit from Plaintiffs in this case, and why OpenAI thinks putting Plaintiffs to the

26   burden of producing this broad swath of information from *Kadrey* would be proportional to

27   OpenAI's needs in *this* case. In other words, simple relevance (if it even is at all relevant, which

28   may very well not be the case) is not the only *sine qua non* of discovery, and the court finds that –

1   without the showing of a more "specific" and "targeted" need that actually relates to *this* case –

2   OpenAI's request for the production of Plaintiffs' responses to interrogatories and requests for

3   admission in *Kadrey* must be **DENIED** as not proportional to the needs of this case.

4   OpenAI also seeks expect witness documents from *Kadrey* on a similar basis – that is, by

5   claiming that they are "also directly relevant to critical aspects of Plaintiffs' theories in this case

6   [because] [f]or example, one key issue that will undoubtedly require expert testimony in both

7   cases is whether Plaintiffs are able to distinguish the harm or apportion[ment] [of] damages

8   allegedly caused by Meta versus OpenAI versus other LLM developers, as is evident from

9   Plaintiffs' *Kadrey* deposition transcripts already produced in this action." *Id*. at 4. OpenAI adds

10   that "Plaintiffs' experts in both cases will almost certainly also opine on core issues common to

11   both cases, such as (a) the traditional or potential markets for Plaintiffs' works and any alleged

12   harm to such markets; (b) whether a market to license books to train LLMs exists; and (c)

13   Plaintiffs' alleged damages, including an assessment of royalties generated by the asserted works

14   common to both cases." *Id*. If that is the case, then it appears that the compelled production of

15   expert witness documents from *Kadrey* would either simply be duplicative of the expert witness

16   documents to be generated in this case, or, that they would merely offer OpenAI differing sets of

17   expert opinions by the experts in this case versus those in *Kadrey* – and that is assuming *arguendo*

18   that OpenAI is correct about the supposedly overwhelming factual and legal overlap between the

19   two cases, something which OpenAI has merely stated, in somewhat conclusory fashion, in the

20   letter brief at bar. In either scenario, however, the usefulness of the *Kadrey* expert witness

21   documents to OpenAI's defense of this case appears to be minimal at best, and therefore the

22   request for their production must similarly be **DENIED** as not proportional to the needs of this

23   case. As Plaintiffs have pointed out, the court has already once denied Plaintiffs' request for

24   cloned discovery from a parallel case involving OpenAI in the Southern District of New York, and

25   OpenAI's request for discovery from the *Kadrey* case – based on the articulated justifications

26   presented here by OpenAI – must be similarly denied as overly broad, marginally relevant (if at

27   all), and not proportionate to the needs of the case.

28   //

1   **IT IS SO ORDERED.**

2   Dated: February 27, 2025

3

4

                         ROBERT M. ILLMAN

5   United States Magistrate Judge

6

7

8

9

10

11

12

13

14

15

16

17

18

19

20

21

22

23

24

25

26

27

28

United States District Court
Northern District of California

10